



Choix éditoriaux de revues scientifiques
pour associer des jeux de données à un
article :
étude d'un échantillon d'articles d'auteurs
du Cirad publiés en 2013

F. Rambour, M .F. Fily
CIRAD
Délégation à l'information
scientifique et technique (Dist)
Montpellier, 18 novembre 2014

Choix éditoriaux de revues scientifiques pour associer des jeux de données à un article : étude d'un échantillon d'articles d'auteurs du Cirad publiés en 2013

----- Résumé

La présente étude tente d'identifier des stratégies ou des pratiques d'éditeurs de revues pour associer des jeux de données aux articles de revues qu'ils publient. Elle porte sur un échantillon d'articles d'auteurs du Cirad publiés en 2013 et référencés dans la base de données institutionnelle Agritrop.

Chaque article sélectionné a été examiné dans ses différents formats (html pour le web, pdf pour la version à imprimer) en vue d'identifier la nature du lien entre l'article et les jeux de données de recherche associés à l'article : données totalement intégrées à l'article et indissociables de l'article ; données individualisées et insérées dans l'article sous forme de fichiers additionnels ; données faisant l'objet d'un article à part entière et publiées dans une revue sous une forme particulière appelée *data paper* ; données déposées par l'auteur (auto-archivées) dans un entrepôt de données indépendamment de l'article mais reliées à cet article et à la revue qui le publie par le biais de liens hypertextes.

Des exemples illustrent chacun de ces modes d'association d'articles et de jeux de données et rendent compte de différents choix éditoriaux de revues scientifiques pour le traitement et la valorisation de jeux de données.

Mots-clés : article de revue, données de la recherche, jeu de données, revue scientifique, entrepôt de données, data paper

Choix éditoriaux de revues scientifiques pour associer des jeux de données à un article : étude d'un échantillon d'articles d'auteurs du Cirad publiés en 2013

----- Sommaire

1. Introduction.....	5
2. Echantillonnage et méthode d'étude.....	5
3. Résultats.....	6
3.1. Données intégrées au corps de l'article	6
3.1.1. Les revues scientifiques du Cirad	6
3.1.2. African Development Review (Wiley)	7
3.2. Jeux de données insérés dans l'article sous forme de fichiers additionnels	8
3.2.1. Theoretical and Applied Genetics (Springer)	9
3.2.2. Carbohydrate Polymers (Elsevier)	10
3.3. Jeux de données publiés sous forme d'article de type <i>data paper</i>	11
3.4. Jeux de données déposés dans un entrepôt et associés à un article par des liens hypertextes	11
3.4.1. PLoS Pathogens (PLoS)	12
3.4.2. Heredity (Nature Publishing Group)	14
3.4.3. BMC Plant Biology (Biomed Central)	15
3.4.4. Cladistics (Wiley-Blackwell)	16
4. Conclusion.....	18

Choix éditoriaux de revues scientifiques pour associer des jeux de données à un article : étude d'un échantillon d'articles d'auteurs du Cirad publiés en 2013

----- Liste des figures

Figure 1 : Figure d'un article de <i>Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux</i>	7
Figure 2 : Tableau d'un article de la revue <i>African Development Review</i> (Wiley)	8
Figure 3 : Fichier additionnel sur le site de la revue <i>Theoretical and Applied Genetics</i> (Springer)	9
Figure 4 : Signalement d'un jeu de donnée dans article au format pdf de la revue <i>Theoretical and Applied Genetics</i> (Springer)	10
Figure 5 : Tableau téléchargeable dans un article de la revue <i>Carbohydrate Polymers</i> (Elsevier)	10
Figure 6 : Fichier additionnel dans un article de la revue <i>Carbohydrate Polymers</i> (Elsevier)	11
Figure 7 : Figure téléchargeable dans un article de la revue <i>PLoS Pathogens</i> (PLoS)	12
Figure 8 : DOI d'une figure dans un article de la revue <i>PLoS Pathogens</i> (PLoS)	13
Figure 9 : Figure d'un article de la revue <i>PLoS Pathogens</i> (PLoS) déposée dans Figshare	13
Figure 10 : Jeu de données d'un article de la revue <i>Heredity</i> (Nature Publishing Group) déposé dans Dryad	14
Figure 11 : Mention d'un jeu de données dans un article de la revue <i>BMC Plant Biology</i> (BioMedCentral)	15
Figure 12 : Accès au jeu de données dans l'entrepôt Array express	16
Figure 13 : Fichier additionnel en <i>open data</i> dans un article de la revue <i>BMC Plant Biology</i> (BioMedCentral)	16
Figure 14 : Rubrique <i>Data compilation</i> dans un article de la revue <i>Cladistics</i> (Wiley-Blackwell)	17
Figure 15 : Jeu de données déposé dans l'entrepôt Morphbank	17
Figure 16 : Jeu de données déposé dans l'entrepôt Dryad	18
Figure 17 : Matrice morphologique sur un site web	18

1. Introduction

La publication des jeux de données de recherche associés à un article scientifique est une pratique que les éditeurs de revues scientifiques tendent à adopter et à développer pour donner une valeur ajoutée aux résultats de recherche décrits dans un article.

L'examen de revues scientifiques révèle une diversité de modalités d'association entre article et jeux de données selon les éditeurs :

- Les données ne sont pas des entités distinctes de l'article mais sont intégrées à l'article avec lequel elles se confondent : l'article et les jeux de données associés ne forment qu'une seule et même entité documentaire ;
- Les données sont insérées dans l'article sous forme de fichiers individuels (par exemple des fichiers *Excel*). L'existence de ces fichiers additionnels est généralement introduite dans le corps de l'article par la mention *supplementary files*. Ceux-ci sont insérés dans le corps de l'article et/ou accessibles isolément sous forme de fichiers distincts sur le site web de la revue ;
- Les jeux de données sont déposés dans une base de données ou entrepôt de données (appelé encore *data repository*) avec des liens hypertextes réciproques entre l'article de la revue et chacun des jeux de données associés à l'article. L'entrepôt de données peut être la propriété de l'éditeur de la revue ou peut être indépendant et externe à l'éditeur ;
- Les données sont publiées sous forme d'articles particuliers appelés *data papers*, décrivant les jeux de données, les circonstances et les méthodes de leur collecte. Les *data papers* peuvent être publiés dans des revues particulières dédiées à ce type de document, les *data journals*, ou dans des revues scientifiques classiques acceptant ce type d'articles.

Afin d'évaluer l'impact de ces modes de publications de jeux de données en association avec des articles, nous avons examiné un échantillon d'articles d'auteurs du Cirad publiés dans différentes revues à comité de lecture en 2013.

Cette étude n'est pas exhaustive. Elle vise à rendre compte, à partir d'exemples de publication empruntés au Cirad, de choix éditoriaux de revues pour la publication de jeux de données ou la mise en relation de jeux de données et d'articles.

2. Echantillonnage et méthode d'étude

A partir de la liste des 706 articles d'auteurs du Cirad publiés dans des revues scientifiques entre janvier et décembre 2013 et enregistrés dans la base institutionnelle Agritrop entre le 25 juillet 2012 et le 14 avril 2014, nous avons procédé à un échantillonnage d'articles publiés dans les revues du Cirad et dans des revues d'éditeurs chez lesquels les auteurs du Cirad publie le plus. Nous n'avons pas distingué les revues librement et gratuitement accessibles (revues en Open Access) des revues payantes pour les lecteurs.

L'échantillon ainsi constitué est composé de 82 articles, hors articles publiés dans les revues du Cirad.

Le corps de chaque article sélectionné a été examiné pour identifier les choix de l'éditeur ou de la revue pour publier les jeux de données associées à un article. Nous avons pu ainsi dresser un panorama de choix éditoriaux de autour des jeux de données.

Pour détecter la présence de jeux de données explicitement associés à l'article, nous avons procédé à une recherche textuelle sur le terme *data* dans les versions web et pdf de l'article. Lorsque cette recherche était infructueuse, nous avons parcouru la page web de l'article pour détecter la présence éventuelle de jeux de données introduits par une autre mention.

La première observation a porté sur la différence de signalement de jeux de données entre la version web et la version pdf d'un article : la version pdf d'un article pouvait par exemple ne mentionner pas

aucun jeu de données tandis que la version web (format html) pouvait indiquer et afficher un lien hypertexte vers un ou plusieurs jeux de données associés à l'article.

3. Résultats

3.1. Données intégrées au corps de l'article

Dans le modèle de revue scientifique traditionnel, les jeux de données associés à un article se confondent avec l'article lui-même : les données apparaissent sous forme textuelle dans le corps de l'article ou sous forme de tableaux ou de graphiques dans l'article. Ceux-ci n'existent pas sous forme de fichiers indépendants : ils ne sont pas affichables et accessibles séparément, ni téléchargeables.

Avantages

L'intégration des jeux de données dans le corps de l'article est totale, ce qui permet de créditer les auteurs de l'article des jeux de données.

Inconvénients

Le format des données doit respecter les contraintes de mise en page de l'article. Les jeux de données ne peuvent être découverts et cités indépendamment de l'article. En outre, le format d'affichage ne facilite pas une extraction et une réutilisation comme un objet à part entière : seul un copier/coller du tableau ou du graphique affichant les données permet de les récupérer.

En outre, si la revue n'est pas en accès libre, les lecteurs non abonnés à la revue ne peuvent pas accéder aux données de l'article.

Dans notre échantillon de 82 articles, les articles contenant uniquement des données intégrées représentent 36 articles, soit 44% de la totalité des articles étudiés.

3.1.1. Les revues scientifiques du Cirad

Les six revues publiées ou soutenues par le Cirad (Bois et Forêts des Tropiques ; Cahiers Agricultures ; Fruits ; Oilseeds and Fats, Crops and Lipids ; Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux ; Natures Sciences Sociétés) ne distinguent pas les jeux de données dans les articles, mais les représentent simplement sous forme de graphiques ou de tableaux confondus au corps de l'article. Les jeux de données ne peuvent pas être cités, ni téléchargés indépendamment de l'article.

Exemple :

Lesnoff M. 2013. *Méthodes d'enquête pour l'estimation des taux démographiques des cheptels de ruminants domestiques tropicaux. Synthèse, limites et perspectives. Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 66 (2) : 57-67.

http://remvt.cirad.fr/cd/derniers_num/2013/REMT13_057_067.pdf

Dans cet article, aucun jeu de données associé n'est cité.

Dans la figure ci-dessous (non téléchargeable), le tableau source cité dans la légende n'est pas accessible non plus.

montre un exemple de décomposition des processus démographiques utilisée récemment pour les enquêtes en milieu tropical (45, 46), et les taux démographiques correspondants sont listés dans le tableau I. D'autres décompositions ont été proposées par le passé (26, 49).

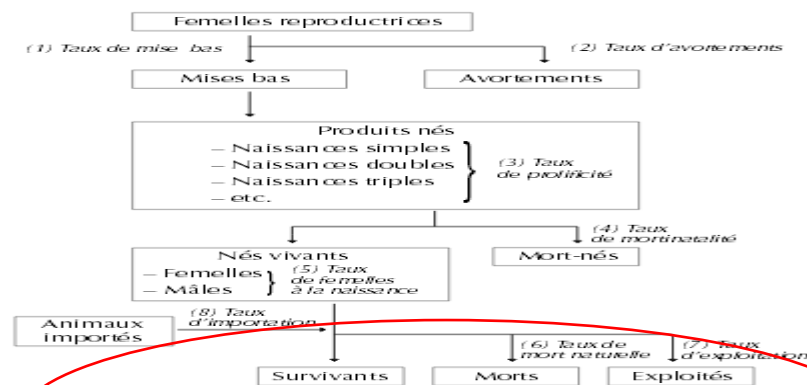


Figure 1 : schéma des événements démographiques dans les troupeaux utilisé pour définir la liste des paramètres démographiques présentés dans le tableau I. Source : Lesnoff et coll., 2007, Quae, adapté de Landais et Sisokho, 1986, Cirad-Hemvt.

Figure 1 : Figure d'un article de *Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux*

3.1.2. African Development Review (Wiley)

Dans l'exemple ci-dessous, les données sont également totalement intégrées au texte de l'article, donc non citables indépendamment et très difficilement réutilisables.

Exemple d'article

Maître d'Hôtel E., Le Cotty T., Jayne T. 2013. Trade policy inconsistency and maize price volatility: An ARCH approach in Kenya. *African Development Review*, 25 (4) : 607-620.

<http://dx.doi.org/10.1111/1467-8268.12055>

3.4 Data

Recherche Google

Many factors are likely to influence food price volatility and could be introduced in X. Basically, food price volatility is related to supply and demands fundamentals, which are likely to include market-specific and broader economic factors (Roache, 2010), and changes in these factors may have large effects because the short-run supply and demand elasticities of food prices are typically low (Balcombe, 2009). The following factors are taken into account in our analysis (see Table 2), and correspond to the ones included in Barrett's model specification (Barrett, 1997).

- *Monthly lagged prices.* There are periods of relatively high and low price volatility, though the underlying unconditional volatility remains unchanged. This principle underlines the choice of an ARCH model. Domestic nominal prices, obtained from the Kenyan Ministry of Agriculture, have been converted into US dollars by the application of an exchange rate (extracted from the Kenyan National Bureau of Statistics).
- *Inflation.* Inflation has an obvious direct effect on food price volatility: to account for this effect, we have been working on Consumer Price Index deflated price series. International prices have been deflated by using the US CPI base 2007 (extracted from the International Labour Organization database) and domestic prices have been deflated using the Kenyan CPI base 2007 (extracted from Kenyan National Bureau of Statistics)
- *Yearly stock levels.* Stocks have an important role in theoretical models of commodity pricing (Williams and Wright, 1991; Deaton and Laroque, 1992). In theory, when stocks are low, volatility is expected to increase; empirical evidence, so far, has been mixed. To account for this potential effect, we used data on Kenyan stock levels from the United State Department of Agriculture database to include the levels of stocks held by the NCPB in our model.
- *Monthly exchange rate.* Volatile exchange rates are likely to induce a higher volatility in food prices, as the riskiness of returns increases (Balcombe, 2009). Exchange rate data, obtained from the Kenyan Bureau of Statistics, are included in the first step of our model.
- *Monthly international prices.* International prices were calculated as prevailing international market prices, extracted from the database of the United Nations Conference on Trade and Development (UNCTAD): we have been working on real prices expressed in US\$/ton base 2007. To do so, the international nominal prices have been deflated by the US CPI.

Table 2. Explanatory variables used

Variable name	Variable description	Sources
Lagged prices	Maize wholesale real prices (ZMK/kg)	Kenyan Ministry of Agriculture
Stock levels	Annual stock level (MT)	USDA PSD database
Exchange rate	ZMK/USD	Kenyan Bureau of Statistics
International price	Maize real international price (USD/kg)	UNCTAD
Policy inconsistency	Import tariffs variability	Ministry of Trade and Industry

Other economic factors that potentially influence food price volatility were not included because of lack of data or redundant information, such as weather patterns,¹ oil price volatility,² and speculation.³ Added to these common economic factors, we integrate in our model one additional factor reflecting policy inconsistency. We characterize policy inconsistency by the dynamic analysis of import tariffs policies in Kenya related to international prices. We generate the variable INCONS as a dummy variable that equals one for the event of simultaneously high international price (higher than its average) and high tariff (higher than its average) or of simultaneously low international prices and low tariff. In all other cases, INCONS equals zero. INCONS accounts for the absence of the expected response to an international price change. When INCONS equals one, policy is inconsistent with the objective of removing effects of international price volatility, and as such is hardly predictable by market operators. More elaborate definitions of inconsistency have been tried, taking into account past values of price in particular, but this very simple measure is sufficient to illustrate our point.

Figure 2 : Tableau d'un article de la revue *African Development Review* (Wiley)

Dans certains articles des revues publiées par l'éditeur Wiley, les figures sont téléchargeables sous forme d'images au format PowerPoint. Elles peuvent donc être utilisées seulement comme une illustration d'une présentation, et non comme un jeu de données de recherche.

Dans tous les articles de l'éditeur Wiley étudiés, les tableaux ne sont jamais téléchargeables.

3.2. Jeux de données insérés dans l'article sous forme de fichiers additionnels

Dans ce modèle éditorial, les données sont présentes dans la version web de l'article sous forme de fichiers additionnels téléchargeables.

Il peut s'agir des tableaux et/ou des figures du texte, partiellement ou totalement téléchargeables (généralement aux formats PowerPoint ou csv) ou de fichiers de données supplémentaires.

Avantages

Le format des données est libéré des contraintes de mise en page de l'article. Toutefois, les formats de publication des jeux de données sont imposés par les éditeurs.

Accessibles directement sur la page web de l'article, les jeux de données sont intégrés dans l'article et peuvent être crédités aux auteurs de l'article

Inconvénients

La taille de ces fichiers de données est souvent limitée à 10 GB.

Les formats des fichiers additionnels de jeux de données et la façon dont ils sont signalés ne répondent pas à des règles uniformes ou standardisées. Les mentions les plus rencontrées sont : *additionnal files*, *supplementary files*, *supplementary material*.

Selon le format de fichier, les données sont plus ou moins facilement réutilisables. Par exemple, un fichier de tableur (csv) est plus facilement réutilisable qu'un tableau disponible dans un format de traitement de texte propriétaire (comme le format doc de Word). Ces fichiers additionnels sont rarement dotés d'un identifiant numérique pérenne (DOI) indépendant de l'article : de ce fait, ils sont difficilement repérables et citables en tant qu'objet scientifique numérique indépendant.

Dans notre échantillon de 82 articles, les articles avec fichiers additionnels représentent 40 articles, soit près de la moitié du corpus étudié.

3.2.1. Theoretical and Applied Genetics (Springer)

Les versions web des articles de cet éditeur ont toutes une structure similaire quelle que soit la revue : page de présentation de l'article puis accès au texte intégral.

Une rubrique *supplementary material* est toujours présente sur la page de présentation. Selon les articles, elle contient ou non des fichiers additionnels.

Exemple de l'article :

Lebeau A., Gouy M., Daunay M.C., Wicker E., Chiroleu F., Prior P., Dintinger J. 2013. Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. *Theoretical and applied genetics*, 126 (1) : 143-158.

<http://dx.doi.org/10.1007/s00122-012-1969-5>

La rubrique *supplementary material* se situe après la rubrique *abstract*, sur la page de présentation. Il n'est donc pas nécessaire d'accéder au texte complet de l'article pour avoir accès aux jeux de données associés.

Dans cet exemple, nous avons accès à des fichiers Word (doc).

Ces fichiers additionnels sont accessibles sur le site web de la revue, même si l'on n'est pas abonné à la revue.

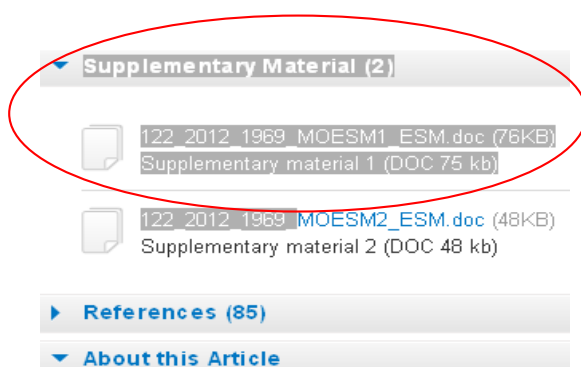


Figure 3 : Fichier additionnel sur le site de la revue *Theoretical and Applied Genetics* (Springer)

Dans la version pdf de l'article, le jeu de données est signalé après la rubrique *abstract*, de la façon suivante :

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1969-5) contains supplementary material, which is available to authorized users.

Figure 4 : Signalement d'un jeu de donnée dans article au format pdf de la revue *Theoretical and Applied Genetics* (Springer)

Le DOI (Digital Object Identifier) indiqué correspond à l'identifiant pérenne de l'article contenant le jeu de données. Il renvoie ainsi à la version web de l'article et non exclusivement et indépendamment au jeu de données associé.

3.2.2. Carbohydrate Polymers (Elsevier)

Exemple d'article :

Rolland-Sabaté A., Sanchez T., Buléon A., Colonna P., Ceballos H., Zhao S.S., Zhang P., Dufour D. 2013. Molecular and supra-molecular structure of waxy starches developed from cassava (*Manihot esculenta* Crantz). *Carbohydrate Polymers*, 92 (2) : 1451-1462.
<http://dx.doi.org/10.1016/j.carbpol.2012.10.048>

Les jeux de données (sous forme de figures, de tableaux...) sont insérés dans l'article, mais sont traités comme des entités indépendantes et sont téléchargeables séparément. Ils sont également accessibles directement sur la page de présentation de l'article, en haut à gauche, dans la rubrique *Figures and tables*.

Chain length distributions of TMS60444 and WAX amylopectins (a and b, respectively) and comparison of chain length distributions of WXA19 and WXB9 amylopectins presented as plots of the differences with TMS60444 (c) and WAXN (d) amylopectins. The gray and the black bars correspond to WXA19 and WXB9 amylopectins, respectively.

Figure options ▼

Table 5.
Chain length distributions of debranched starches obtained from HPAEC-PAD.

Type of Starch	Peak DP		% distribution					Average CL	Highest detectable DP
	I	II	DP 6–9	DP 6–12	DP 13–24	DP 25–36	DP ≥ 37		
TMS60444	11	43	11.6 (0.20)	30.9 (0.54)	48.0 (0.27)	12.5 (0.18)	8.6 (0.63)	19.1 (0.29)	79
WXA19	11	43	11.3 (0.03)	30.2 (0.01)	48.2 (0.01)	12.4 (0.08)	9.2 (0.06)	19.3 (0.02)	82
WXB9	11	43	11.6 (0.02)	30.4 (0.12)	47.9 (0.25)	12.3 (0.03)	9.4 (0.16)	19.4 (0.03)	81
WAXN	12	44	9.3 (0.17)	28.8 (0.61)	49.0 (0.22)	13.0 (0.68)	9.2 (0.20)	19.5 (0.04)	83

Standard deviations are given in parenthesis.

Table options ▲

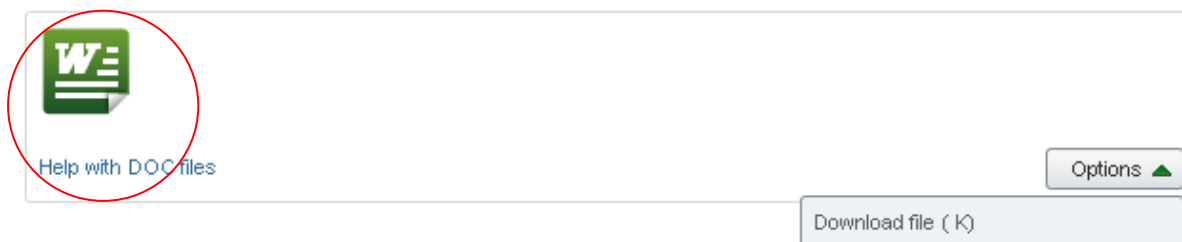
Branch chain length distributions obtained by HPAEC-PAD agreed well with those determined by A4F-MALLS, even if $\overline{DP}_w/\overline{DP}_n$ numerical values were slightly different.

View in workspace
Download as CSV

Figure 5 : Tableau téléchargeable dans un article de la revue *Carbohydrate Polymers* (Elsevier)

D'autres données non incluses dans le texte sont également signalées comme *supplementary data* et accessibles en cliquant sur le logo illustrant le logiciel de traitement de texte Word.

amylopectins (Supplementary data S2) were very close, thus confirming their very similar branching structure, despite the slight differences observed in terms of average branching degree.



3.5.3. Branched chain length distribution

Figure 6 : Fichier additionnel dans un article de la revue *Carbohydrate Polymers* (Elsevier)

3.3. Jeux de données publiés sous forme d'article de type *data paper*

Certains articles (appelés *data papers*) sont dédiés à la description de jeux de données ou de collections de jeux de données. Ils peuvent être publiés dans des revues particulières appelées *data journals* ou dans des revues classiques acceptant les *data papers*. Les *data journals* peuvent fournir eux-mêmes un espace de dépôt des données ou préconiser aux auteurs d'articles des entrepôts de données externes pour le dépôt de leurs jeux de données.

Le *data paper* ne concerne que la description du jeu de données et de ses métadonnées mais ne fournit pas une analyse de résultats.

Avantages

Les jeux de données sont décrits de manière standardisée et sont mis à disposition dans des formats connus et pérennes.

Les auteurs de l'article sont crédités des jeux de données décrits dans le *data paper*.

Chaque jeu de données est doté d'un identifiant pérenne, généralement un DOI.

La citation et la réutilisation de ces jeux de données sont aisées.

Des liens hypertextes réciproques sont effectués entre article et jeu(x) de données associé(s), et leur pérennité est assurée.

Les entrepôts assurent la conservation pérenne des données qui y sont déposées.

Inconvénients

Ce type d'article est encore très peu utilisé.

A ce jour, les principaux domaines pour lesquels il existe des *data journals* sont l'écologie, la biodiversité et les sciences naturelles, le domaine biomédical, l'environnement, les géosciences et l'archéologie.

Dans notre échantillon, aucun des 82 articles n'est un *data paper*.

3.4. Jeux de données déposés dans un entrepôt et associés à un article par des liens hypertextes

Les jeux de données étayant les résultats décrits dans un article scientifique peuvent être déposés dans des entrepôts de données reconnus (appartenant aux éditeurs scientifiques ou externes) avec liens réciproques entre l'article scientifique et les données.

Certains éditeurs, comme *Nature Publishing Group (NPG)* ou *PLoS*, font de la mise à disposition des données dans un entrepôt une condition préalable à l'acceptation de l'article.

Ces entrepôts de données peuvent être généralistes (par exemple les entrepôts Dryad et Figshare), disciplinaires (par exemple GenBank), ou dans certains cas institutionnels.

A ce jour, les entrepôts disciplinaires n'existent que dans quelques disciplines : biologie, sciences de la vie, sciences du sol, chimie.

Avantages

Les jeux de données sont décrits de manière standardisée et mis à disposition dans des formats connus et pérennes.

Les entrepôts assurent la conservation pérenne des données déposées.

Il n'existe souvent pas de restriction sur le volume des données déposées.

La pérennité des liens réciproques avec l'article peut être assurée si le dépôt dans l'entrepôt a été effectué avant ou au moment de la soumission de l'article à la revue.

Inconvénients

Le dépôt dans certains entrepôts est payant, ce qui augmente le coût de publication de l'article.

Si le dépôt des données a été effectué après la publication de l'article, le lien vers le jeu de données peut ne pas être mentionné sur l'article. Le jeu de données est alors « invisible » et sa consultation, sa réutilisation et sa citation sont problématiques.

De nombreux éditeurs permettent à la fois la publication de données en tant que matériel supplémentaire associé à l'article et le dépôt de données dans un entrepôt. Selon les éditeurs, il peut s'agir des mêmes données ou de données différentes.

Dans notre échantillon de 82 articles, les jeux de données déposés dans un entrepôt représentent 5 articles, soit 6% de la totalité des articles étudiés.

3.4.1. PLoS Pathogens (PLoS)

Exemple d'article :

Berri F., Rimmelzwaan G.F., Hanss M., Albina E., Foucault-Grunenwald M.L., Lê V.B., Vogelzang-van Trierum S.E., Gil P., Camerer E., Martinez D., Lina B., Lijnen R., Carmeliet P., Riteau B. 2013. Plasminogen controls inflammation and pathogenesis of influenza virus infections via fibrinolysis. *PLoS Pathogens*, 9 (3) : e100322 (12 p.).

<http://dx.doi.org/10.1371/journal.ppat.1003229>

Dans cet exemple, l'article propose des données directement téléchargeables à partir de la page web de l'article et l'accès à d'autres données déposées dans un entrepôt.

Chaque figure est identifiée par un DOI et peut être téléchargée, directement à partir de la version web de l'article au format PowerPoint, ou aux formats d'images png ou tiff.

Chaque figure peut être ainsi citée indépendamment de l'article.

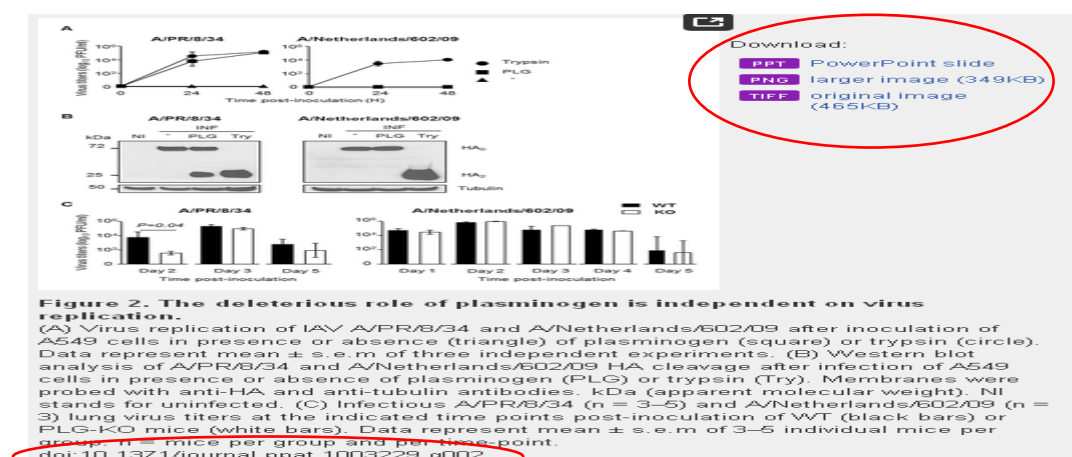


Figure 7 : Figure téléchargeable dans un article de la revue *PLoS Pathogens* (PLoS)

Dans la version pdf de l'article, la légende de chaque figure indique son DOI.

Ce DOI renvoie directement sur la figure téléchargeable, sur la version web de l'article.

Figure 2. The deleterious role of plasminogen in
 Netherlands/602/09 after inoculation of A549 cells in pres
 \pm s.e.m of three independent experiments. (B) Western bl
 in presence or absence of plasminogen (PLG) or trypsin
 molecular weight). NI stands for uninfected. (C) Infectious
 points post-inoculation of WT (black bars) or PLG-KO mice
 group and per time-point.

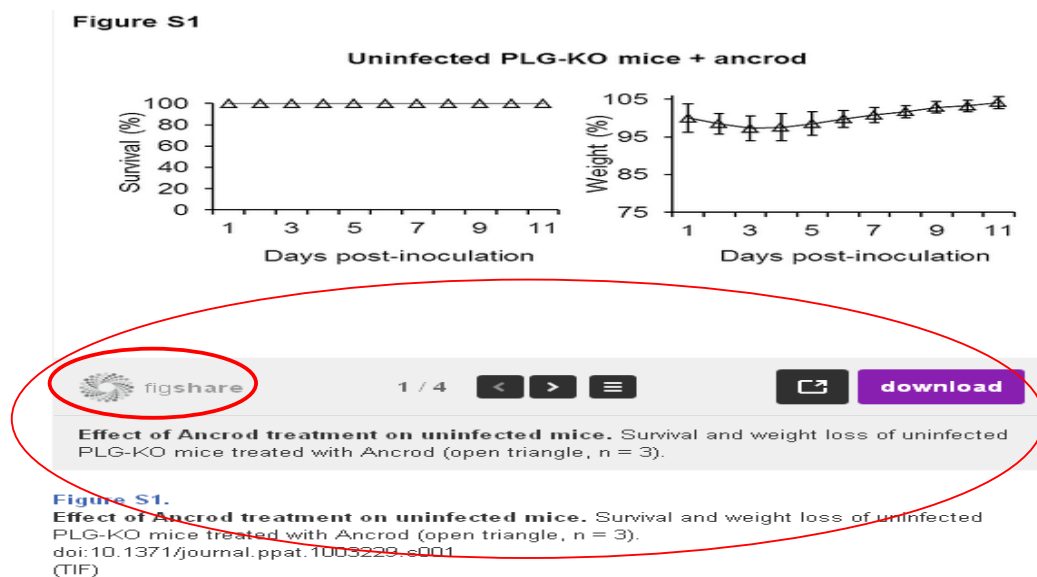
doi:10.1371/journal.ppat.1003229.g002

Figure 8 : DOI d'une figure dans un article de la revue *PLoS Pathogens* (PLOS)

En fin d'article, juste avant la rubrique des références bibliographiques, une rubrique intitulée *Supporting Information* contient plusieurs figures téléchargeables, soit directement via le bouton *Download* présent sur la version web de l'article, soit à partir du site web de l'entrepôt Figshare dans lequel ont été déposées les données.

Chaque figure accessible et téléchargeable a son propre DOI.

Les figures déposées sur Figshare sont en libre accès, sous licence CC-by qui représente l'obligation de citer la source en cas de reproduction ou de réutilisation.



Après le clic

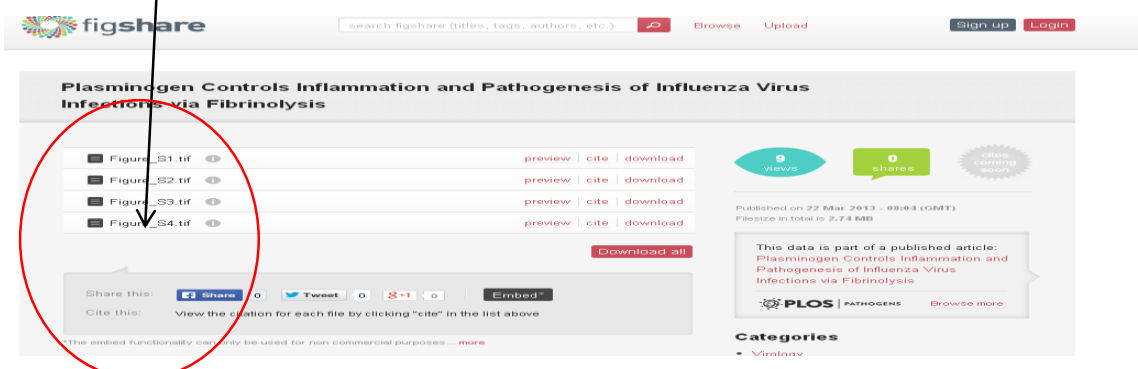


Figure 9 : Figure d'un article de la revue *PLoS Pathogens* (PLOS) déposée dans Figshare

3.4.2. Heredity (Nature Publishing Group)

L'éditeur *Nature Publishing Group* a défini une obligation de dépôt des données dans des entrepôts disciplinaires seulement pour certains types de données standardisées comme les séquences protéiques, les séquences ADN, les structures macromoléculaires, les données cristallographiques, etc.

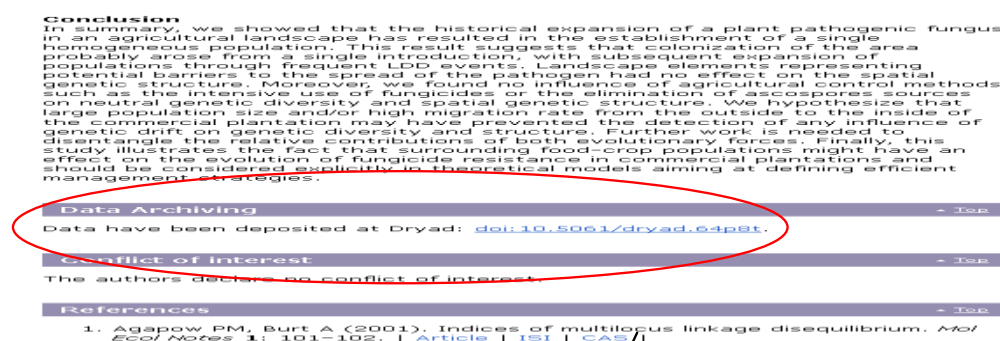
Pour les autres jeux de données, l'éditeur recommande aux auteurs d'articles, le dépôt des jeux de données associés dans un entrepôt de données (cf. *Availability of data and materials*, <http://www.nature.com/authors/policies/availability.html>)

Par exemple, les chercheurs proposant des articles dans le domaine des sciences doivent s'assurer, par l'intermédiaire d'une *checklist*, qu'ils donnent suffisamment de détails méthodologiques sur leurs analyses et qu'ils précisent les modalités d'accès à leurs données (cf. *Nature Reporting Checklist For Life Sciences Articles*, may 2013, <http://www.nature.com/authors/policies/checklist.pdf>).

Exemple d'article :

Rieux A., De Lapeyre de Bellaire L., Zapater M.F., Ravigné V., Carlier J. 2013. Recent range expansion and agricultural landscape heterogeneity have only minimal effect on the spatial genetic structure of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Heredity*, 110 (1) : 29-38.
<http://dx.doi.org/10.1038/hdy.2012.55>

La mention du jeu de données figure dans une rubrique intitulée *Data Archiving* située en bas de l'article, avant la rubrique *References* de la bibliographie.



Le jeu de données est identifié par un DOI: [10.5061/dryad.64p8t](https://doi.org/10.5061/dryad.64p8t). Si nous cliquons dessus, nous arrivons sur la page de description et de téléchargement du jeu de données dans l'entrepôt *Dryad*.

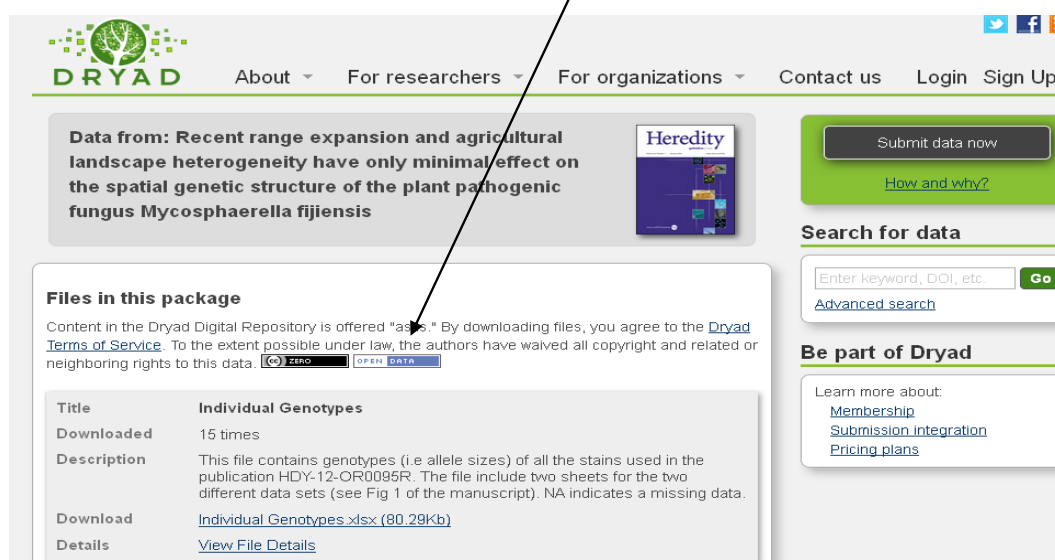


Figure 10 : Jeu de données d'un article de la revue *Heredity* (Nature Publishing Group) déposé dans Dryad

Les données déposées dans l'entrepôt Dryad sont transférées dans le domaine public, sous licence CC0 et peuvent être réutilisées sans autorisation.

Dryad suggère toutefois de citer l'article d'où sont issues les données ainsi que le jeu de données lui-même.

3.4.3. BMC Plant Biology (Biomed Central)

L'éditeur *Biomed Central* exige que les données décrites dans les articles, incluant les données brutes pertinentes, soient rendues librement accessibles. Ce peut être sous forme de fichiers supplémentaires (*additional files*) accessibles sur la page de présentation de l'article, ou, plus fréquemment, de jeux de données déposés dans des entrepôts.

Exemple d'article

Germanà M.A., Aleza P., Carrera E., Chen C., Chiancone B., Costantino G., Dambier D., Deng X., Federici C.T., Froelicher Y., and al. 2013. Cytological and molecular characterization of three gametoclones of *Citrus clementina* : Interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. *BMC Plant Biology*, 13 (1) : 8 p.
<http://dx.doi.org/10.1186/1471-2229-13-129>

La présence d'un jeu de données est mentionnée en fin d'article, avant la rubrique *References*. Il est indiqué qu'une demande de dépôt de ce jeu de données a été effectuée auprès de l'entrepôt *ArrayExpress*.



Figure 11 : Mention d'un jeu de données dans un article de la revue *BMC Plant Biology* (BiomedCentral)

En interrogeant l'entrepôt *ArrayExpress* avec le numéro d'identification *E-MEXP-3926* indiqué dans l'article, les fichiers de données brutes sont retrouvés et peuvent être téléchargés.

Home Experiments Arrays Submit Help About ArrayExpress Feedback Login

ArrayExpress > Experiments > E-MEXP-3926

E-MEXP-3926 - Comparative genomic hybridization by array of three gametoclonal lines of *Citrus clementina*

Status	Released on 12 November 2013, last updated on 3 June 2014
Organism	<i>Citrus clementina</i>
Samples (6)	Click for detailed sample information and links to data
Array (1)	A-MEXP-1017 - CFP-Citrus_20k
Protocols (4)	Click for detailed protocol information
Description	Gene deletion analysis of three independent gametoclonal lines of <i>Citrus clementina</i> by comparative genomic hybridization
Experiment types	comparative genomic hybridization by array, comparative genome hybridization, dye swap design, strain or line design
Contact	Gabino Rios <gribs@ivia.es>
Citation	Cytological and molecular characterization of three gametoclonal lines of <i>Citrus clementina</i> . Germana MA, Aleza P, Carrera E, Chen C, Chiancone B, Costantino G, Dambier D, Deng X, Federici CT, Froelicher Y, Guo W, Ibáñez V, Juárez J, Kwok K, Luro F, Machado MA, Naranjo MA, Navarro L, Ollitrault P, Rios G, Roose ML, Talon M, Xu Q, Gmitter FG Jr. <i>BMC Plant Biology</i> 13(1):129 (2013), Europe PMC 24020638
MIAME	Platforms Protocols Factors Processed Raw
Files	Investigation description Sample and data relationship Raw data (1) Array design Click to browse all available files
Links	Send E-MEXP-3926 data to GENOME SPACE

Figure 12 : Accès au jeu de données dans l'entrepôt Array express

Dans ce même article, est également signalé un *additional file* en *open data* (données ouvertes librement réutilisables), qui contient des figures au format PowerPoint, directement téléchargeables à partir de la version web de l'article.

Additional file 1: Figure S1. Two homozygous plants of *Citrus clementina* Hort. ex Tan., cv. Nules from France (FRA) and Italy (ITA). The third, from Spain (ESP), was described in Aleza et al. [22]. **Figure S2.** Flow cytometry analyses of DNA content for the plants from France (FRA) and Italy (ITA).
 Format: PPTX Size: 453KB [Download file](#)

OPEN DATA

Figure 13 : Fichier additionnel en *open data* dans un article de la revue *BMC Plant Biology* (BiomedCentral)

En septembre 2013, l'éditeur Biomed Central a décidé que ces fichiers additionnels seraient placés par défaut en accès libre (*open access*) sous licence *Creative Commons* CC0, les plaçant donc dans le domaine public (logo *open data*). Ces jeux de données peuvent ainsi être réutilisés sans autorisation et sans obligation d'en citer la source.

3.4.4. Cladistics (Wiley-Blackwell)

Exemple d'article

Heraty J.M., Burks R.A., Cruaud A., Gibson G.A.P., Liljeblad J., Munro J., Rasplus J.Y., Delvare G., Jansta P., Gumovsky A., Huber J.T., Woolley J.B., Krogmann L., Heydon S., Polaszek A., Schmidt S., Darling D.C., Gates M.W., Mottern J., Murray E., Dal Molin A., Triapitsyn S., Baur H., Pinto J.D., Van Noort S., George J., Yoder M. 2013. A phylogenetic analysis of the megadiverse Chalcidoidea (Hymenoptera). *Cladistics*, 29 (5) : 466-542.
<http://dx.doi.org/10.1111/cla.12006>

La dernière rubrique de l'article, intitulée *Data compilation*, décrit la façon dont les jeux de données associés à l'article, ont été constitués, et affiche le lien hypertexte vers les fichiers de données déposés dans les entrepôts.

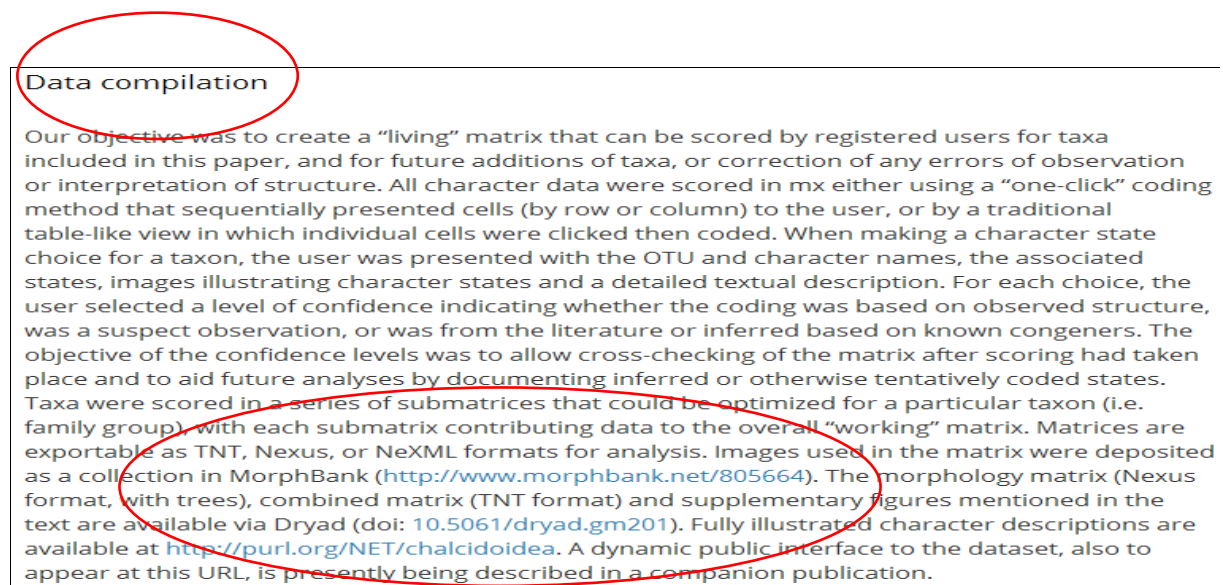


Figure 14 : Rubrique *Data compilation* dans un article de la revue *Cladistics* (Wiley-Blackwell)

Le premier lien (<http://www.morphbank.net/805664>) mène dans l'entrepôt *MorphBank* où sont accessibles toutes les matrices exportables en formats tnt, nexus ou neXML.



Figure 15 : Jeu de données déposé dans l'entrepôt Morphbank

Le second lien ([http://dx.doi.org/10.5061/dryad.gm201](https://dx.doi.org/10.5061/dryad.gm201)) pointe sur l'entrepôt *Dryad* où l'on retrouve le détail de chaque « Figure » de l'article et l'accès à d'autres informations.

DRYAD About ▾ For researchers ▾ For organizations ▾

Data from: A phylogenetic analysis of the megadiverse Chalcidoidea (Hymenoptera)

Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data. [CC0](#) [OPEN DATA](#)

Title	Chalcid_datasetCombined 2012
Downloaded	40 times
Download	Chalcid_datasetCombined.TNT (1.246Mb)
Details	View File Details

Title	Chalcidoidea Morphology Matrix 2012
Downloaded	56 times
Description	Contains the original characters with all trees for mapping.
Download	Nexus File ChalcidMorphology withtrees Mesquite.nex (183.0Kb)
Details	View File Details

Figure 16 : Jeu de données déposé dans l'entrepôt Dryad

Le troisième lien (<http://purl.org/NET/chalcidoidea>) ouvre un site web identifié par une adresse web pérenne ou PURL (Persistent uniform resource locator).

Ce site donne accès à toutes les données de la matrice morphologique de *Chalcidoidea*.

Chalcidoidea Morphology Matrix

Home
Characters
About

Introduction

A Morphological Matrix for Chalcidoidea

Chalcidoidea is one of the most megadiverse groups of insects. Ranging from the minute and morphologically bizarre male of *Dicopomorpha echmepterygis* (Mymaridae), the smallest known specimen of which is 0.13 mm long, to the veritable giants such as females of *Leptofoenus* (Pteromalidae), which exceed 20 mm, there are estimated to be more than 500,000 species in this single superfamily, of which only about 22,000 species have been described (Heraty, 2009; Noyes, 2011). Their morphological diversity is staggering (Fig. 1). Males of *D. echmepterygis* have lost eyes, ocelli, mouthparts, antennal flagellum, wings, tarsi except for a highly modified arolium, and virtually any other feature that place them as parasitic wasps. *Galearia latreillei* (Eucharitidae) has a grotesquely enlarged scutellum (Fig. 1h) similar to some Perilampidae, whereas males of Agaoninae may be reduced to turtle-like fighting machines, morphologically completely unlike their corresponding females, and females of *Cameronella* (Pteromalidae) with dart-shaped ovipositor sheaths are barely comparable to the distantly related aphelinid wasps, some of which attack the same host family, Eriococcidae. Convergent morphology is rampant, and enlarged femora, enlarged acropleura, reductions in the number of antennal and tarsal segments, and reductions in wing venation, wings and other features have all been proposed as being independently derived in very distantly related taxonomic groups. This extreme numerical and morphological diversity has resulted in a large number of higher taxa being described relative to other superfamilies of parasitic Hymenoptera, with 19 families and 85 subfamilies currently being recognized, and this has prevented any single individual from being able to conduct a comprehensive phylogenetic analysis based on morphology.

For any questions about the characters contact [CONTACT US](#)

Back to the top

Supported by the National Science Foundation (NSF) grant EP-0341149 DEB-0730616, and the informatics team at [Biology File](#) (INHS). All opinions are that of the researchers and not the NSF. [POWERED BY](#)

Figure 17 : Matrice morphologique sur un site web

4. Conclusion

L'étude conduite sur un échantillon d'articles scientifiques d'auteurs du Cirad, publiés en 2013, permet de découvrir et de caractériser différents choix éditoriaux de revues scientifiques pour associer les jeux de données aux articles qu'elles publient. Cette pratique éditoriale, qui tend à se généraliser parmi les revues, contribue à valoriser les résultats de la recherche tout en donnant une valeur ajoutée aux articles eux-mêmes.

Près de la moitié des revues dans lesquelles les chercheurs du Cirad publient (44% dans l'échantillon étudié) ne donnent néanmoins pas accès aux jeux de données de l'article. C'est le cas notamment des six revues produites ou publiées par le Cirad : Bois et Forêts des Tropiques ; Cahiers

Agricultures ; Fruits ; Oilseeds and Fats, Crops and Lipides ; Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux ; Natures Sciences Sociétés.

Lorsque des jeux de données sont associés aux articles, il s'agit dans la moitié des articles étudiés, de tableaux, de graphiques ou d'images accessibles sur le site de la revue sous forme de fichiers additionnels à l'article et téléchargeables. Ces fichiers ne sont que très rarement identifiés et traités comme des objets numériques autonomes et pérennes.

Très peu d'articles (6% dans l'échantillon étudié) donnent accès à des jeux de données déposés dans des entrepôts qualifiés, à même d'assurer la pérennité des données et de leur accès. Toutefois, cette pratique devrait se développer, certains éditeurs scientifiques rendant obligatoire le dépôt dans un entrepôt, de(s) jeu(x) de données associés à l'article soumis pour publication.

Aucun *data paper* n'a été identifié parmi les articles d'auteurs du Cirad publiés en 2013 : ce type de publication est récent et encore peu connu de la communauté scientifique. Il existe d'autre part très peu de *data journals* dans les domaines de recherche du Cirad.

La présente étude met également en évidence l'absence de norme dans l'affichage, la localisation et le référencement de jeux de données associés à un article de revue. Les mentions utilisées pour indiquer la présence d'un jeu de données et son accès dans un article sont hétérogènes. La localisation du jeu de données dans l'article ou la page web qui l'affiche n'est pas toujours facilement identifiable, les formats de fichiers des données acceptés sont différents d'une revue à l'autre, les entrepôts préconisés ou imposés par les revues sont propres à chaque revue.